

VVenC 기반 MPEG MIV 프레임워크 검증

*오관정 정준영 이광순 서정일

한국전자통신연구원

*kjoh@etri.re.kr

Verification Test on VVenC based MPEG MIV Framework

*Oh, Kwan-Jung Jeong, Jun Young Lee, Gwangsoon Seo, Jeongil

Electronics and Telecommunications Research Institute

요약

인공 지능, 사물 인터넷, 클라우드 컴퓨팅, 빅데이터, 모바일을 중심으로 시작된 4차 산업혁명은 최근 메타버스(Metaverse) 시대로 자연스럽게 확장되어 가고 있다. 메타버스 시대의 핵심 기술인 VR(Virtual Reality)/AR(Augmented Reality)/MR(Mixed Reality)과 이를 디스플레이 하기 위한 HMD(Head Mounted Display)나 글래스(glass) 장치들은 사용자에게 고품질 영상뿐만 아니라 3차원 공간에 대한 6DoF(Degrees of Freedom)급의 자유도를 제공해야 한다.

이에 발맞춰 MPEG에서는 최근 실사 및 CG(Computer Graphics)로 제작된 다시점 텍스처 및 깊이 영상에 대한 효과적인 압축 방법을 찾기 위한 MIV(MPEG Immersive Video) 표준화를 진행 중이다. 본 논문은 최신 2차원 압축 표준인 VVC(Versatile Video Coding)를 기반으로 MPEG MIV 1.0 표준 프레임워크를 검증해본다. 특히 높은 압축율에 비해 복잡도가 높은 VVC 표준의 특성을 고려하여 HHI에서 개발한 최적화 코덱인 VVenC를 이용하여 압축 효율과 복잡도 관점에서 MIV 표준 활용을 위한 최적의 솔루션을 찾고자 한다.

1. 서론

멀티미디어에 대한 대표적인 국제 표준화 기구인 MPEG(Moving Picture Experts Group)에서는 2016년 MPEG-I(Immersive)라는 이름으로 사용자에게 6DoF(Degrees of Freedom)의 몰입형 미디어 서비스를 제공하기 위해 필요한 비디오, 오디오, 그래픽스, 시스템 기술에 대한 표준화를 위한 프로젝트를 시작했다. 그 중 비디오 파트 관련해서 MPEG에서는 MIV(MPEG Immersive Video)[1]라는 이름으로 단계적으로 표준화를 진행하고 있으며, 현재는 3DoF+에 대한 표준인 MIV 1.0이 완료 단계에 이르렀다. MIV 1.0 표준은 DIBR(Depth Image Based Rendering)[2] 기술을 기반으로 운동 시차(motion parallax)가 있는 몰입형 비디오를 제공하기 위해 여러 시점의 텍스처 영상과 깊이 영상을 부호화한다.

MIV는 2001년 3DAV(3D Audio Visual)이란 이름으로 시작되어, 이후 FTV(Free-viewpoint TeleVision)으로 계승된 MPEG의 3차원 영상 표준의 3세대 표준이다. 2006년 7월에 시작되어 2009년 3월에 완료된 1세대 FTV 표준인 MVC(Multi-view Video Coding)의 경우 다시점 영상을 부호화함으로써 사용자에게 보다 넓은 시역을 제공하고자 했다. 이때 MVC는 송신단의 입력 시점과 수신단의 출력 시점의 수가 같았고, 다시점 텍스처 영상만을 다뤘다. 2007년 4월에 시작해서 2016년 6월에 마무리된 2세대 표준인 3DV(3D Video)는 넓은 시역에 추가적으로 자연스러운 중간 시점을 제공하기 위해 MVC와 달리 다시점 텍스처와 깊이 영상을 다뤘다. 그리고 표준의 범위는 아니지만 깊이 영상 예측과 DIBR 기술을 이용하여 중간 시점 영상을 합성하는 렌더링을 위한 참조

소프트웨어도 함께 개발했다. MVC와 3DV 표준 모두 다시점 영상을 효율적으로 압축하기 위해 기존의 2차원 영상 압축 표준인 H.264/AVC(Advanced Video Coding)나 H.265/HEVC(High Efficiency Video Coding)[3]에 시점간 예측(interview prediction)을 확장한 프레임워크를 기반으로, 2차원 영상 코덱의 부호화 기술들을 최적화했다.

앞서 소개한 MPEG의 기존 3차원 영상 표준인 MVC나 3DV와 달리 MIV 1.0 표준의 가장 큰 특징은 2차원 비디오 표준과의 호환성을 유지한다는 점이다. 이는 MIV 1.0이 참조 표준인 V3C(Visual Volumetric Video-based Coding)의 여러개의 패치(patch)로 구성된 아틀라스(atlas)를 이용하는 프레임워크를 따르기 때문이다. 따라서 MIV는 이를 위한 전처리, 후처리, 그리고 메타데이터 기술에 대한 표준화이다. MIV에서 다루는 데이터는 이전 3DV 표준과 동일하게 다시점 텍스처와 깊이 영상이지만 보다 다양한 카메라 배치를 허용함으로써 표현 가능한 3차원 공간에 대한 자유도가 훨씬 넓어졌다.

현재 MIV 1.0 표준은 HEVC와 최신 영상 코덱인 VVC(Versatile Video Coding)[4]를 기반 코덱으로 활용하고 있다. 본 논문에서는 그 중 VVC를 기반으로 MIV 1.0 표준 프레임워크의 성능을 검증해본다. 검증 실험에는 고성능 부호화 효율에도 불구하고 높은 복잡도를 보이는 VVC의 참조 소프트웨어 코덱인 VTM(VVC Test Model)[5] 대신에 Fraunhofer HHI에서 개발한 VVC의 최적화 코덱인 VVenC(Versatile Video enCoder)[6]와 VVdeC(Versatile Video deCoder)[7]의 활용 가능성을 검토하고자, 다양한 구성(configuration) 방식을 이용해 실험을 수행하고, 그 결과를 비교 분석한다.

2. MPEG MIV 프레임워크

MIV는 현재 TMIV(Test Model for Immersive Video)라는 참조 소프트웨어를 통해 검증이 가능하다. 아래 그림 1과 2는 MIV의 인코더와 디코더 구조를 보여주고 있다. 인코더 입력은 다시점 텍스처와 깊이 영상 그리고 시점별 카메라 파라미터이다. MIV 인코더는 해당 데이터를 재구성해 텍스처와 깊이 영상에 대한 아틀라스 영상을 만들고, 이를 표현하기 위한 부가 정보들을 메타데이터로 정의한다. 각 아틀라스 영상들은 기존의 2차원 비디오 표준 코덱을 이용해 부호화하는데, 현재 MIV 1.0은 HEVC와 VVC를 활용하고 있다.

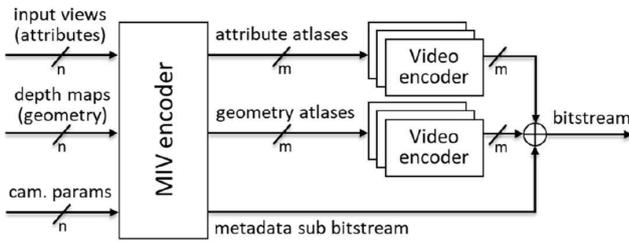


그림 1 MIV 인코더

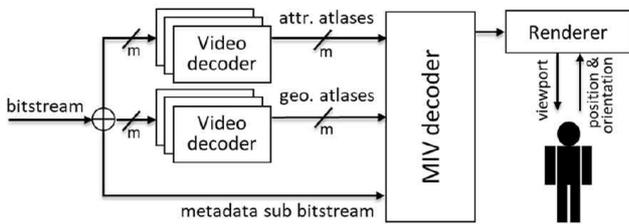


그림 2 MIV 디코더

특히, VVC의 경우는 VTM 대신 VVenC라는 HHI의 최적화 코덱을 활용한다. VVenC는 5가지 종류의 구성 모드(Slower, Slow, Medium, Fast, Faster)를 제공하고 있는데 현재 MIV의 CTC(Common Test Conditions)는 Slow 모드를 권장한다. 그러나 서비스 환경에 따라 대역폭과 인코더 및 디코더의 복잡도에 대한 조건이 상이하다. 따라서 다양한 부호화 조건에 대한 실험을 통해 VVenC 기반의 MIV 서비스에 대한 최적의 솔루션을 찾고자 한다.

3. 실험 결과

검증을 위해 MIV의 실험 영상 중 “Kitchen”, “Chess”, “Fan”, “Carpark” 영상을 이용해 17 프레임에 대한 부호화 실험을 수행했다. 실험 조건은 MIV의 CTC[8] 조건을 따랐다. 본 실험에서 anchor는 MIV와 동일하게 Slow 조건을 이용했다. 표 1은 5개의 실험 조건에 대해 전방위 비디오의 객관적 품질 평가 척도인 WS-PSNR(Weighted-to-Spherically-uniform PSNR) 기준으로 고비트율과 저비트율 환경에서의 효율을 비교하여 보여주고 있다. 그림 3은 표 1의 실험 영상별 비트율-왜곡 곡선을 보여주고 있다. 그림에서 보듯이 “Slower”, “Slow”, “Medium”은 유사한 결과를 보였고, “Fast”와 “Faster”는 상대적으로 부호화 효율 측면에서 큰 차이를 보였다.

표 1 BD-rate 성능 비교 (Slow 조건 대비)

	Slower	Medium	Fast	Faster
Fan	-7.3% -6.9%	6.7% 6.9%	112.0% 90.2%	166.1% 128.1%
Kitchen	-4.5% -6.2%	5.1% 5.3%	29.1% 30.9%	58.2% 62.7%
Carpark	-5.2% -5.0%	5.6% 6.0%	74.9% 71.2%	186.2% 158.4%
Chess	-7.6% -8.0%	11.9% 11.7%	53.5% 52.5%	143.5% 141.6%
평균	-6.2% -6.5%	7.3% 7.5%	67.4% 61.2%	138.5% 122.7%

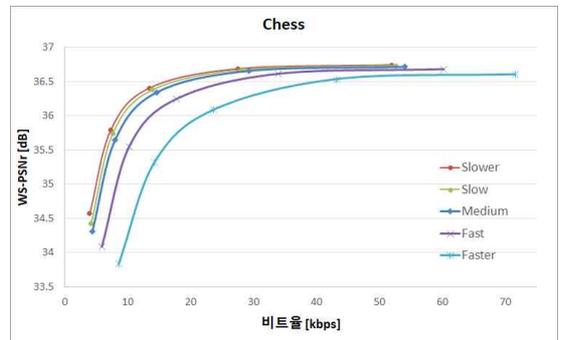
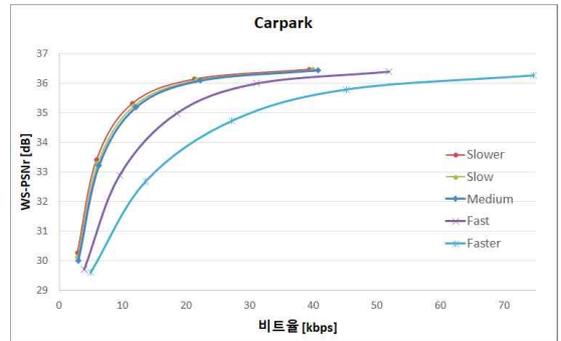
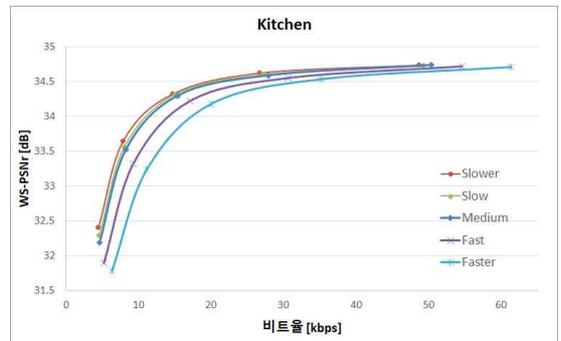
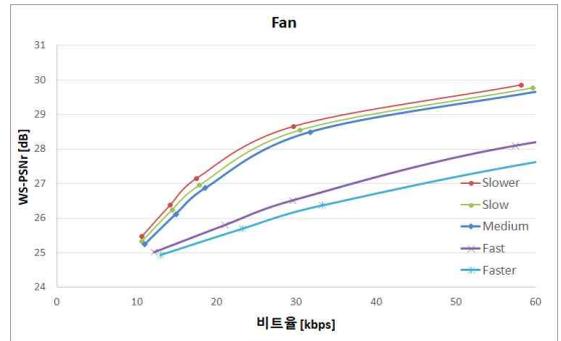


그림 3 비트율-왜곡 곡선

표 2는 “Slow” 조건대비 상대적인 인코딩 시간을 비교하여 보여주고 있다. 실험 결과 Slower 조건은 인코딩 시간이 4배 정도 증가하고 추가 성능 향상은 6% 내외에 머물렀다. Faster 조건은 실험 영상에 따라 인코딩 및 디코딩 시간이 최대 90% 정도 줄었지만, Kitchen 영상을 제외하고는 비트율이 1.5배 가까이 증가했다. 반면에 Medium과 Fast 조건이 성능과 복잡도 측면에서 활용 가능한 성능을 보여줬다. 그림 4는 부호화 성능 대비 인코딩 시간을 그래프로 보여주고 있다. 그림에서 보듯이 Medium 조건이 부호화 성능대비 인코딩 시간면에서 상대적으로 우수함을 확인할 수 있었다.

표 2 인코딩 시간 비교 (Slow 조건 대비)

	Slower	Medium	Fast	Faster
Fan	392.0%	34.1%	13.0%	9.2%
Kitchen	384.1%	40.9%	15.4%	10.7%
Carpark	486.9%	42.4%	21.3%	16.6%
Chess	472.0%	39.9%	17.1%	11.8%
평균	433.8%	39.3%	16.7%	12.1%

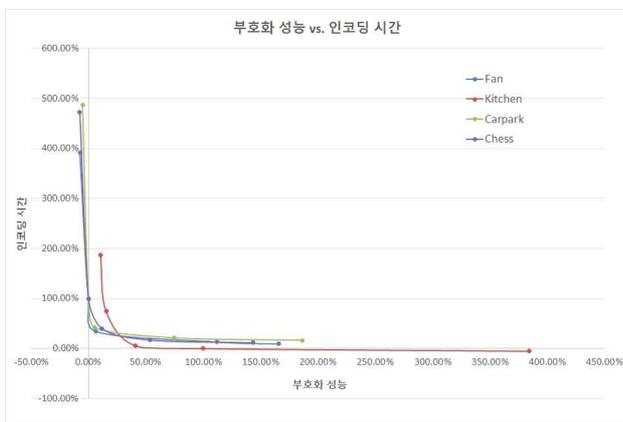


그림 4 부호화 성능 대비 인코딩 시간 비교

4. 결론

본 논문에서는 최신 영상 압축 표준인 VVC의 최적화된 참조 코덱인 VVenC를 기반으로 다양한 조건에 대한 MPEG MIV의 성능과 인코더 및 디코더 복잡도에 대한 검증 실험을 수행했다. 실험 결과 Medium 조건이 부호화 성능 및 복잡도 측면에서 MIV 서비스를 위한 현실적인 코덱으로 판단된다. 그러나 이는 2차원 영상 압축 코덱인 VVC의 기존 실험 조건에 따른 것으로, VVC의 다양한 부호화 기술들이 MIV에 친화적인지에 대한 추가 검증 및 최적화를 통해 추가적인 성능 개선 및 복잡도 향상이 가능하리라 생각된다.

감사의 글

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No. 2018-0-00207, 이머시브전문연구실)

참고문헌

- [1] J. M. Boyce et al., "MPEG Immersive Video Coding Standard," in Proceedings of the IEEE, 2021.
- [2] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," in Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems, vol. 5, pp. 93-104, San Jose, USA, January 2004.
- [3] ITU-T and ISO/IEC JTC 1, High Efficiency Video Coding, Rec. ITU-T H.265 and ISO/IEC 23008-2 (HEVC), Apr. 2013 (and subsequent editions)
- [4] ITU-T and ISO/IEC JTC 1, Versatile Video Coding, Rec. ITU-T H.266 and ISO/IEC 23090-3(VVC), July 2020.
- [5] VTM software repository, version VTM-12.0. Available online: https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM
- [6] VVenC software repository. Available online: <https://github.com/fraunhoferhhi/vvenc>
- [7] VVdeC software repository. Available online: <https://github.com/fraunhoferhhi/vvdec>
- [8] ISO/IEC JTC 1/SC 29/WG 04 MPEG Video Coding, "Common Test Conditions for MPEG Immersive Video," N0051, October 2020.